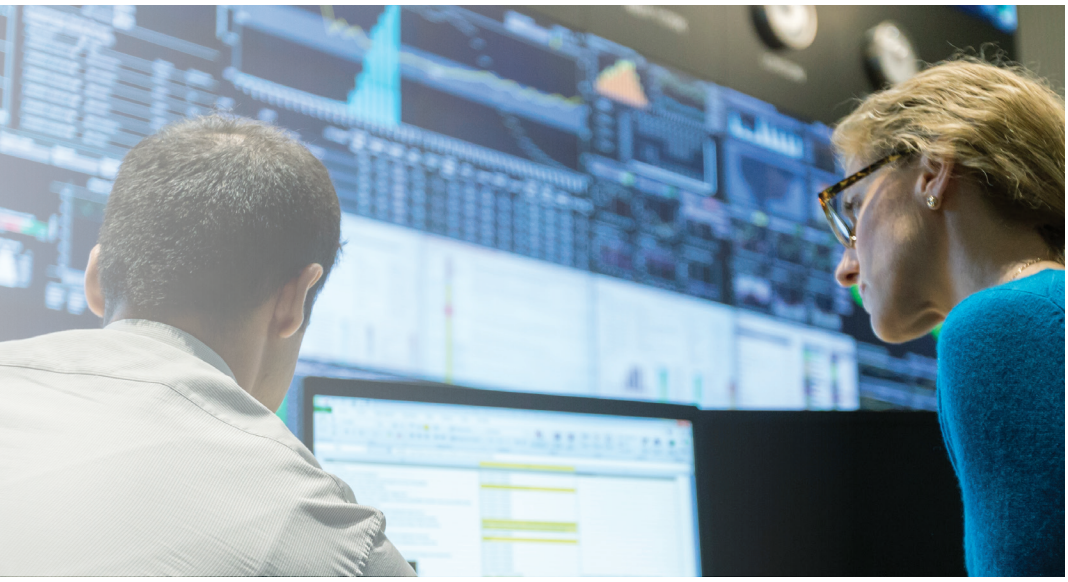
OPTUM®

The great data transition

Making the move from claims to EHR data



For more than a decade, life sciences companies have relied on claims data as their go-to source for health economics and outcomes research (HEOR) studies, and for commercial projects such as patient journey studies. While the utilization of these data is certainly valuable, working in a claims-only environment greatly limits the understanding of patient types and important differences in treatments and outcomes.

Many researchers and analysts have been reluctant to incorporate electronic health record (EHR) data into their real-world data strategies. But as the pressure to prove the value of existing and new therapies increases, it's critical for life sciences firms to leverage additional and more comprehensive data assets. EHR data is a proven source for gaining additional clinical perspective beyond claims and data sourced from highly controlled clinical trials.

Understanding patients with EHR data

In the past, life sciences companies were primarily focused on small molecules in major disease classes, such as hypertension, diabetes and asthma. Anonymized patient-level claims data became a key data source because it helped determine treatment lines, concomitant therapy usage, and outcomes such as hospitalizations and associated costs. While claims data provides value, it does not offer a complete understanding of patient types, symptoms, disease progression and treatment rationale.

This has become more important as life science companies have shifted to higher cost treatments, usually in complex and often more-targeted disease states. Furthermore, it is difficult to quantify the general overall health of patients just by looking at claims data. Conditions may or may not be coded. Regular measures of progression, lab values and key observations from physicians and patient complaints are generally unknown. This is why claims data has often been supplemented with chart studies and primary market research to answer the "whys" associated with the "what was billed" information provided by claims. However, these other tools are also limited. Charts can provide simple measures, such as vital signs, pain scores and physician notes on treatment response, but collecting charts in an unbiased manner is difficult and costly. There is often artificiality to primary market research where practitioners reference best practices that differ from actual real-world behavior. Also, due to the focused nature of primary market research, only a select number of questions can be addressed in a single study.

Despite the limitations of claims data, some organizations have been slow to embrace new data forms like EHR in their real-world evidence strategy development.

If EHR data can be so valuable, and everyone now knows about it, why has its adoption not been faster and broader?

- **EMR system structure.** The structure and quality of data provided by early EMR vendors were often crude. Data elements such as diagnoses, procedures and vital signs housed in structured fields may have been insufficient or too superficial for research use. Less structured data, particularly free-form text in notes, are a key feature of EMR systems, but rudimentary systems limit its value. Beyond ensuring HIPAA compliance, there was no processing to extract and create meaningful structure from the notes. Early forays into medical records came up short and left traditional claims users where they started, after considerable investments of time and money.
- **Resistance to EHR data validation.** Experiences with other data types, such as claims data, may have led to unreasonable expectations and evaluation criteria for EHR data. Claims data is well defined because it is driven by payers and their reimbursement requirements. EHR data is captured by the physician, for the physician. What is most important during a patient visit, and what is likely to be pertinent or relevant to their current or future care, gets recorded. Because EHR entries are much less driven by universal code sets, there is resistance from those who consider EHR data validation requirements not strict enough. As a result, the valuable data in the EHR is often overlooked.
- **Belief that EHR data is "incomplete."** The initial reaction to EHR data is that it is incomplete and of poor quality. However, this reaction is actually part of the very reason why transitioning to EHR data is valuable. The "incomplete" nature of EHR data truly represents real-world conditions. Users of EHR data should look at it from the perspective of what information is gained, as opposed to what is not populated or easily understood. If everything that a physician entered is made available, then in effect, there is no missing data, only desired data not recorded. The strict quantification required for significance testing in clinical trials does largely not exist in real-world practice. Each patient is a trial of its own, and the appropriateness of outcome measures can vary patient by patient based upon general health, patient choices and other factors.

EMR vs. EHR

Electronic medical record (EMR) and electronic health record (EHR) may sound the same, but there are distinct differences. The **EMR** is the digital capture of a patient's medical chart with one provider. It is a single point-of-care record captured by one EMR system. It contains the patient's basic medical history, diagnosis and treatment, but is not shareable with other providers.

The **EHR** is the complete health record of a patient across all providers and treatment settings, captured from multiple EMR systems. It offers a comprehensive, research-ready view of the patient's complete health care journey and is shareable across providers and facilities.

- **Unfamiliarity with "real-worldliness" of EHR data.** Physicians do not need a strict quantification of disease severity or progression. Broad categories that they can quickly recognize based upon more limited measures and perceptual impressions are sufficient. The gut feeling that a doctor has of a patient may be hard to document, but experience has convinced them that it can successfully guide a patient's treatment. The lesson learned is not just to look for what you want in EHR data, but to also try to give meaning to what you find. With increasing frequency of mandated treatment protocols within integrated delivery networks (IDNs) and more requirements for external reporting (such as CMS quality reporting requirements), the uniformity and consistency of EHR records has improved. However, learning to use what you find in the EHR and putting yourself into the head of the physician interacting with the patient is good guidance for utilizing and understanding the data.

Many concerns around EHR data continue to be addressed. Vendors capture more complete patient experiences, primarily by sourcing data from all sites of care within the increasing number of IDNs. These networks provide the patient a more integrated care experience, and when records from different EMRs within the systems are integrated, the records can be highly valuable to researchers.

The implementation and growing sophistication of natural language processing (NLP) has helped address the quality and usability of less structured/unstructured data in EMRs. More experience with researchers' desired outputs and the ability of NLP to extract words and phrases that accurately represent important clinical concepts can result in specificity and sensitivity that may rival what is produced by human coders.

Adapting how we think about EHR data

How can we help these new users utilize EHR data — and do so successfully? First, we must help them understand that maximizing success is an incremental process. Many users expect to hit the ground running with their new EHR data, but this expectation should be tempered.

Learnings from the information in an EHR record can be placed on a continuum from the evident to the emergent.

The "evident" data are those elements that come from more structured fields and have little ambiguity around them. Straightforward mentions in a note may fall into this category. Some of these elements are code-driven (diagnoses and procedures) and others, such as vital signs and standard test results, have well-established measurements and recording norms. While these elements still need to be checked before use, the time to value for applying these data should be minimal. The degree to which these readily usable elements can be found will be market dependent. Some therapeutic areas have lots of well-populated key elements that can quickly enhance what was learned from claims alone, while others require more exploration of what is routinely recorded in the data.

The "emergent" data require a more creative process for finding and using what is available. Much of the emergent information in EHR data comes from physician notes. NLP can help find something as simple as a term indicating a side effect with an associated measure of severity (for example, intense pain), or as complex as implementing emerging definitions of disease progression and future event risk. High-quality NLP will populate tables that are often the building blocks for an analysis, but they are rarely a finished product. Essentially, we are making the qualitative *quantitative* and easier to retrieve and use. NLP can often find associations in data to guide analysis, but it requires the judgment of the researcher to create reasonable analyses.

Given the evident and emergent information in EHR data, there will be a lot of variance in the time required to find usable records for a study. Whereas attrition tables from claims are usually easy to create, the same attrition tables created from EHR often require considerably more time. Points at which records are eliminated may require further thought with regard to what inclusion/exclusion criteria can be utilized. Researchers need to attend to the absolute number of remaining patients with sufficient information without too much focus on starting-population sizes. Also, given that many patients will not have all data elements of interest, creating composite measures that utilize what is available without dropping a patient because some data is not available is very important. Whereas claims patients tend to be included/excluded from studies based upon the length of the observation period available for the patient, EHR record inclusion is more driven by having enough of the needed information on the record. With time, expertise will develop on how to maximize the number of records that can be included without sacrificing study quality.

Integrated claims and EHR data can be the holy grail for insightful analytics.

For EHR data to be considered of high quality and usable, it must have the following attributes:

- **Be inclusive of all patient care delivery episodes.** While single point-of-care EMR data may be useful for the study of acute and episodic treatments, researchers generally require more than this. An EHR offering that picks up only limited portions of a patient's entire health journey based upon payer, site or EMR system utilized is likely to have niche utility at best. In fact, we believe the term EHR should be reserved for records meeting a minimum completeness standard.
- **Fidelity to source.** What is passed on to the user should be what is really in the source EMR, without distortions, omissions or errors. EHR providers should have the ability to combine EMR data across multiple systems populating a target database accurately by mapping the source EMR data. Through a staging process with quality assurance and quality control measures, we can be convinced that the target EHR system contains appropriately combined EMR data.
- **Ability to accurately extract key information from provider notes.** If information recorded in free-form notes is not provided, EHR records are of minimal value. Providing raw notes does not allow efficient knowledge extraction across large numbers of patients. Techniques and technology like NLP must be deployed to create structured data from free-form text fields. This is a time-intensive activity and one size does not necessarily fit all. EHR providers should work with a client to create meaningful custom concept extractions from prescriber notes in a transparent way.

Progressing on the EHR data learning curve

While many organizations have made substantial progress in getting up to speed with EHR data, true industry leaders and experts are still relatively small in number. To a large extent, the EHR user community is gaining capabilities as the data and collective skill sets improve.

Data providers must be willing to commit adequate resources to help users gain knowledge with their data. While vendors may not have the specific therapeutic area knowledge that end-user researchers have, they should have learnings from prior client engagements that are generalizable. Understanding where new users are coming from and helping traditional claims users understand how to access and analyze EHR data appropriately can help avoid initial user frustration. To optimize the journey of discovery, users should tap into multiple resources.

- The best way to gain insights is for organizations to share knowledge with each other.
- To accelerate knowledge acquisition, smaller organizations, or those with limited commitment to real-world data projects, should leverage external resources, such as conferences and seminars.
- Regulatory agencies like the FDA have recognized the potential impact of real-world evidence (RWE) and are working with organizations to define RWE requirements for usable studies. The FDA recognizes that RWE can come from diverse sources, including EHR, medical claims, product and disease registries, lab results and cutting-edge technology paired with consumer mobile devices.

Beyond EHR, integrated data offers the ultimate value

Integrating diverse data sources yields the greatest rewards for life sciences organizations. Claims can show a patient's adherence to treatments that are ordered (particularly retail prescriptions), and the cost burden of disease occurrence and progression to a health system. This financial component is vital to health economics researchers who need to turn care choices into an economic and clinical argument. Marrying these claims elements with the clinical details and specificity found only in EMRs can provide a compelling story. In addition, overlapping elements in claims can provide confirmatory and sometimes expanded evidence of the extent of delivered care. Given that EMRs are site-specific, and claims flow to processors and payers not typically receiving EMR data, it rests on data integrators to find a way to obtain a critical mass of integrated data for analysis.

While claims and EHR data are an obvious first choice for integration, many other data sources can be integrated with EHR data records. The internet of things (IoT) is greatly impacting the extent to which devices can exchange data in the doctor's office or hospital. Researchers would love to have such items as imaging data, full EKG/ECG, surgical monitoring and many other detailed results attached to the patient record in the EMR. Currently, we rely on the physician to make a note of pertinent information seen in these other sources, but it is often highly limited. Given the reductions in costs of storage and the increased ability to seamlessly exchange data, it is a question of when, not if, such enhanced records might be available.

As the health care system and technology evolve, EMR has the potential to grow in value.

Device data from diabetes pumps and pacemakers can also be integrated with EMR-sourced data as part of a health record. Fitness wearables that capture respiration, heart rhythms and sleep patterns can enrich a patient record and explain factors influencing health. Patient-reported outcome data that can be regularly entered into apps is yet another emerging source of data for EMR integration.

While integrating more patient data can present challenges for maintaining HIPAA compliance and honoring patient privacy requests, there likely will be many incentives for patients to opt-in to letting health care providers see more of their physiological functioning.

There's no time like the present

Now is the time for most pharmaceutical, biotech and medical device companies to endorse EHR data as a necessary component of their data holdings. This may mean bringing data for one or more therapeutic areas in-house for the first time, or broadening the use of data that may already be licensed across departments and functions. As health care systems and technology evolve, EHR data has the potential to grow immensely in value. What we consider good EHR today will likely be augmented with additional information. We have already seen EHR data quality increasing over recent years as practices consolidate, EMR systems improve and physicians recognize the importance of EHR data beyond their own use.

EHR data isn't an easy answer for all research questions that arise, and traditional data sources such as claims will continue to be useful. However, given the impact of immediately accessible information, as well as the emergent learnings that will come from progressing up the learning curve with EHR data, no competitive organization will want to be in the position where it has to rapidly make up for lost time by sitting this one out.

Support from peers, vendors, consultants and other industry users can all help to facilitate this EHR-based journey toward better understanding of the clinical patient.

Optum can help you make the move to EHR data. Connect with us today.

Email: connected@optum.com
Phone: 1-866-306-1321
Visit: optum.com/lifesciences

optum.com/lifesciences

11000 Optum Circle, Eden Prairie, MN 55344